REPORTE DE INVESTIGACIÓN

1. Nombre del profesor

Dra. Lucía A. Ruiz Galindo

2. Proyectos registrados ante Consejo Divisional

607 Análisis Multivariado y de Series de Tiempo y

891 Modelos con fundamentos microeconómicos

3. Líneas de generación y/o aplicación de conocimiento

Econometría, Series de Tiempo**,** Modelos micro y macroeconómicos

4. Área o Grupo de Investigación

Grupo de Investigación de Modelación Económica Teórica y Aplicada
(en proceso de aprobación)

ECONOMETRÍA APLICADA USANDO R

NORMALIDAD

Por

LUCÍA A. RUIZ GALINDO

UNIVERSIDAD AUTÓNOMA METROPOLITANA
DIVISIÓN DE CIENCIAS SOCIALES Y HUMANIDADES
DEPARTAMENTO DE ECONOMÍA
GRUPO DE INVESTIGACIÓN MODELACIÓN ECONÓMICA
TEÓRICA Y APLICADA

Diciembre, 2015.

Normalidad

Lucía A. Ruiz Galindo[1]

1. Introducción

En la especificación del modelo de regresión lineal se distinguen dos partes, una determinística y otra estocástica. Una vez estimado el modelo y habiendo aprobado la evaluación económica del mismo, se llevan a cabo pruebas de especificación correcta, es decir, pruebas mediante las cuáles se verifican los supuestos del modelo, los de su parte determinista, la que corresponde a la combinación lineal de los parámetros, y los de la estocástica, la asociada al término aleatorio.

De manera más específica, las pruebas de especificación correcta consisten en estudiar si la información empírica incorporada en el modelo, la que se utiliza para su estimación, proporciona evidencia a favor o en contra de los supuestos tanto de la parte determinista del modelo como de la aleatoria o estocástica. En la primera, generalmente se estudia si las variables independientes son las únicas que explican a la dependiente, si hay permanencia estructural en los parámetros y si la forma funcional en que se han introducido las variables es correcta o no, entre otras, mientras que en el término estocástico, se analizan los supuestos Gauss-Markov, que establecen que los errores aleatorios tienen media cero, son independiente de las variables explicativas, homoscedásticos y no autocorrelacionados, y también se verifica que se satisfaga el supuesto de normalidad, todo ello se lleva a cabo usando pruebas de hipótesis estadísticas. En este contexto, es importante indicar que cuando se realiza la prueba de un supuesto particular, se asume que todos los demás se satisfacen (supuesto *ceteris paribus*).

El objetivo de este Capítulo es doble: estudiar la importancia e implicaciones del supuesto de normalidad en el modelo de regresión lineal y de manera específica en la inferencia estadística de sus parámetros, y presentar en R, aplicaciones de la prueba de Jarque-Bera (Jarque-Bera 1980, 1987), utilizada para detectar si los términos estocásticos en el modelo siguen o no una distribución normal.

[1] Departamento de Economía, Universidad Autónoma Metropolitana - Azcapotzalco.
laruizg@correo.azc.uam.mx

En la segunda Sección de este Capítulo se hace una breve presentación del modelo de regresión lineal, en la tercera se desarrollan dos procedimientos de estimación: el de mínimos cuadrados ordinarios y el de máxima verosimilitud y se analizan brevemente las propiedades de los estimadores resultantes, a partir de ellos se estudia la importancia que tiene el supuesto de normalidad de los errores estocásticos en la inferencia estadística y de manera más precisa, en la formulación de intervalos de confianza y de pruebas de hipótesis para todos los parámetros del modelo de regresión lineal, en la cuarta Sección se formula la prueba de Jarque-Bera para analizar si los errores satisfacen el supuesto de normalidad, utilizando para ello los residuos como *proxis* de los errores o términos estocásticos, en la quinta Sección se presenta la forma en que se realiza esta prueba en R y se muestran algunas aplicaciones de la misma, en la sexta se exponen las causas e implicaciones que tendría el hecho de que el supuesto de normalidad no se satisfaga y además, se muestran posibles soluciones, finalmente, en la séptima Sección, se plantean algunas conclusiones.

2. Modelo general de regresión lineal

2.1 Especificación del modelo

Considere que la variable dependiente es explicada por K-1 variables independientes, esto es,

$$y_t = \beta_1 + \beta_2 x_{t2} + \ . \ . \ . + \beta_K x_{tK} + \varepsilon_t \qquad (1)$$

donde $\beta_1, \ . \ . \ . \ , \beta_K$ son los parámetros del modelo, y_t es la variable dependiente, las x_{tk}'s, $k = 2, \ . \ . \ . , K$, son las variables independientes, ε_t es el término o error estocástico, *t*, $t = 1, \ . \ . \ . \ , T$, es un índice que indica el número de la observación y T es el total de observaciones.

El modelo está formulado en el momento o periodo *t*, por ello las variables y el término estocástico están indexados con ese subíndice; mientras que el subíndice *k* en las variables independientes o explicativas, indica el número de la variable en la ecuación de regresión. Por ejemplo, x_{t5} y x_{tK}, señalan la variable 5 y la *K*, ambas en el momento *t* mientras que x_{5k} y x_{100k}, indican la observación 5 y 100 de la variable *k*.[2]

En la especificación anterior se distinguen dos partes, la determinista o también conocida como forma funcional, dada por

$$\beta_1 + \beta_2 x_{t2} + \ldots + \beta_K x_{tK}$$

y la estocástica, que no es más que el término o error aleatorio ε_t. En la parte determinista los parámetros deben de plantearse en forma lineal de manera que el modelo sea lineal en ellos; por su parte, las variables dependiente e independientes, aunque introducidas de manera lineal, pueden no serlo. Debe hacerse notar que de acuerdo a la especificación anterior, los parámetros no cambian al paso del tiempo, no tienen subíndice *t*, por ello se dice que hay permanencia estructural o que no hay cambio estructural.

El modelo en (1) se puede formular de manera matricial como sigue

$$y = X\beta + \varepsilon, \tag{2}$$

donde $y = (y_1, y_2, \ldots, y_T)'$,

$$X = \begin{pmatrix} 1 & x_{12} & x_{13} & \ldots & x_{1K} \\ 1 & x_{22} & x_{23} & \ldots & x_{2K} \\ 1 & x_{32} & x_{33} & \ldots & x_{3K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{T2} & x_{T3} & \ldots & x_{TK} \end{pmatrix},$$

$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)'$ y $\beta = (\beta_1, \beta_2, \ldots, \beta_K)'$. Observe que el vector *y* está constituido por las T observaciones de la variables dependiente, la matriz X de dimensión $T \times K$, por una columna de unos asociada al término independiente y las *K*-1 columnas restantes corresponden a las observaciones de las variables independientes, el vector β de dimensión *K*, por los parámetros del modelo y ε por los T términos estocásticos, uno por cada periodo.

2.2 Supuestos de la forma funcional.

S1. Linealidad en los parámetros.

[2] Esta especificación y todo lo que sigue es válido cuando en lugar de variables en series de tiempo se introducen en corte transversal.

S2. Las *K*-1 variables independientes son las únicas que explican a la dependiente.

S3. El número de observaciones *T*, es mucho mayor que el de parámetros *K*.

S4. Las variables explicativas son linealmente independientes de manera que ninguna es combinación lineal de otra o de otras y por tanto el rango de *X* es *K*.

S5. Los parámetros no cambian en la muestra, es decir, hay permanencia estructural.

2.3 Supuestos Gauss-Markov

Los supuestos Gauss Markov son sobre el término estocástico.[3]

SGM1. $E(\varepsilon_t) = 0,\ \forall t = 1,\ .\ .\ .\ .\ , T.$

SGM2. $\{\varepsilon_1, \varepsilon_2, .\ .\ .\ , \varepsilon_T\}$ y $\{x_{1k}, x_{2k}, .\ .\ .\ , x_{Tk}\}$ son independientes $\forall k = 2,\ .\ .\ .\ .\ , K.$

SGM3. $V(\varepsilon_t) = \sigma^2,\ \forall t = 1,\ .\ .\ .\ .\ , T.$

SGM4. $Cov(\varepsilon_t, \varepsilon_s) = 0,\ \forall t, s = 1,\ .\ .\ .\ .\ , T,\ t \neq s.$

SGM5. ε_t se distribuye Normal, $\forall t = 1,\ .\ .\ .\ .\ , T.$

Los supuestos SGM1, SGM3-SGM5 establecen que los términos estocásticos son elegidos de manera no correlacionada de una distribución normal con media y varianza constante, esto último debido a que ellos son homoscedásticos (SGM3). En notación matricial esas condiciones se pueden formular como

$$\varepsilon \sim N(\mathbf{0}, \sigma^2 I),$$

donde $\mathbf{0}$ e I son de manera respectiva, un vector de ceros de dimensión T y la matriz identidad de $T \times T$.

3. Importancia de la distribución normal en la inferencia estadística

En esta Sección se hace una exposición sucinta de como el supuesto de normalidad de los términos estocásticos es utilizado en la inferencia estadística del modelo, es decir, en la estimación puntual de sus parámetros, en el planteamiento de intervalos de confianza y en la formulación de pruebas de hipótesis.

---

[3] En todo el documento, los momentos poblacionales y todas las distribuciones están condicionados a la información disponible de las variables en el modelo.

La estimación puntual de los parámetros suele realizarse mediante el método de mínimos cuadrados ordinarios (MCO) y el de máxima verosimilitud (MV), pero el primero no utiliza el supuesto de normalidad, mientras que en el de MV es fundamental. Obtenidos los estimadores, con el propósito de plantear intervalos de confianza y hacer pruebas de hipótesis, es necesario determinar las distribuciones de esos estimadores y como estos dependen de los errores estocásticos, sus distribuciones estarán determinadas por la normalidad. A continuación se presentan los aspectos básicos de la inferencia estadística del modelo de regresión lineal.

3.1 Estimación puntual de los parámetros

Una vez especificado el modelo de regresión lineal, se estiman los K parámetros en la ecuación: $\beta_1,\ \beta_2, .\ .\ ., \beta_K$, y el asociado a la varianza del término estocástico: σ^2, de manera que el total de parámetros que se deben estimar es K+1. Los métodos mediante los que se estima el modelo son el de mínimos cuadrados ordinarios (MCO) y el de máxima verosimilitud (MV).

El método de MCO consiste en minimizar la suma de cuadrados de los errores estocásticos, es decir,

$$\min S(\beta_1, \beta_2, .\ .\ ., \beta_K) = \sum_{t=1}^{T} \varepsilon_t^2$$

o equivalentemente en forma matricial,

$$\min S(\beta) = \varepsilon' \varepsilon = (Y - X\beta)'(Y - X\beta).$$

Resolver este problema implica plantear las condiciones de primer orden o ecuaciones normales a partir de las cuáles se determina el punto crítico

$$\hat{\beta}_{MCO} = (X'X)^{-1} X'Y$$

y mediante la matriz de segundas derivadas, el hessiano, se analiza que efectivamente en él se alcanza un mínimo. El procedimiento de MCO sólo proporciona el estimador de las betas, no el de la varianza de los errores, σ^2, pero se propone como estimador mínimo cuadrático de la varianza el siguiente

$$\hat{\sigma}_{MCO}^2 = \frac{1}{T-K}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 = \frac{1}{T}\hat{\varepsilon}'\hat{\varepsilon}$$

donde $\hat{\varepsilon}_t$ son los residuos, es decir,

$$\hat{\varepsilon}_t = y_t - \hat{y}_t, \quad t = 1, \dots, T.$$

y

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_{t2} + \ . \ . \ . + \hat{\beta}_K x_{tK}$$

es el ajuste del modelo o también nombrado el ajuste del modelo..

Por su parte, el método de MV como su nombre lo indica, maximiza la función de verosimilitud de los errores o de manera equivalente, su logaritmo, para lo cual se debe de considerar que $\varepsilon_1, \varepsilon_2, . \ . \ . , \varepsilon_T$ constituyen una muestra aleatoria, es decir, un conjunto de variables aleatoria independientes e idénticamente distribuidas. Como puede observarse, a diferencia del procedimiento de MCO en el que no se hace ningún supuesto sobre la distribución de los errores estocásticos para obtener los estimadores de los parámetros, el de MV debe considerar una distribución de los mismos y por supuesto, esa es la normal.

De esta forma, la función de verosimilitud de los términos estocásticos considera que son seleccionados de manera independiente de una distribución normal y dados los supuestos SGM1, SGM3-SGM5, su media debe ser cero, y su varianza σ^2, esto es,

$$\varepsilon_t \sim N(0, \sigma^2).$$

Por ello, la función de verosimilitud es

$$L(\beta_1, \ldots, \beta_K;\ \varepsilon_1, \ldots, \varepsilon_T) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{T}{2}} e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2}$$

y la log- verosimilitud está dada por

$$l(\beta_1, \ldots, \beta_K;\ \varepsilon_1, \ldots, \varepsilon_T) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2$$

Nuevamente, obtener la solución de maximizar esta función, implica determinar las condiciones de primer orden y solucionar el sistema de ecuaciones para obtener los puntos críticos, que en este caso están dados por

$$\hat{\beta}_{MV} = (X'X)^{-1}X'Y$$

y

$$\hat{\sigma}_{MV}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 = \frac{1}{T}\hat{\varepsilon}'\hat{\varepsilon}.$$

Observe que $\hat{\beta}_{MCO} = \hat{\beta}_{MV}$, pero $\hat{\sigma}_{MCO}^2 \neq \hat{\sigma}_{MV}^2$ los cuáles difieren en sus grados de libertad y por ende, en sus correspondientes propiedades.

El teorema de Gauss-Markov establece que los estimadores mínimo cuadráticos, los de las betas, son los mejores estimadores lineales e insesgados, MELI o BLUE por sus siglas en inglés, *best linear unbaised estimator*, es decir, dentro de los lineales e insesgados son los de mínima varianza. Como $\hat{\beta}_{MCO} = \hat{\beta}_{MV}$, entonces los máximos verosímiles también son MELI.

Por su parte, el estimador de MCO de la varianza es insesgado, pero su varianza es mayor que la correspondiente al estimador máximo verosímil, pero éste a pesar de ser más eficiente que el mínimo cuadrático, es sesgado. De aquí en adelante y dado que el estimador mínimo cuadrático de beta es igual al máximo verosímil, se nombrara simplemente beta gorro, es decir, $\hat{\beta}_k, k = 1, \ldots, K$ o bien, en su forma vectorial $\hat{\beta}$.

Como puede observarse, el supuesto de normalidad del error aleatorio es de suma importancia para obtener los estimadores máxima verosímiles de los parámetros del modelo de regresión, no así para los mínimos cuadráticos, que prescinde de ese supuesto, y también es útil para determinar distribuciones que adquieren relevancia al formular intervalos de confianza y hacer pruebas de hipótesis para los parámetros del modelo incluyendo la σ^2, tal y como se verá en las siguientes Secciones.

3.2 Intervalos de confianza y pruebas de hipótesis

Implicaciones inmediatas del supuesto de normalidad de los errores estocásticos son las que se tienen sobre la distribución de la cantidad pivotal a partir de la cual se plantean los intervalos de confianza, y la del estadístico para llevar a cabo las pruebas de hipótesis, tanto para los parámetros en la especificación del modelo, las β_k, como para la varianza del error aleatorio, σ^2.

El desarrollo de esos dos tipos de inferencia para las betas generalmente se hace bajo dos escenarios, cuando σ^2 es conocida y cuando no lo es, pero independientemente de ello se parte del hecho de que los términos estocásticos constituyen una muestra aleatoria, es decir, son variables aleatorias independientes e idénticamente distribuidas, esto es,

$$\varepsilon_t \sim N(0, \sigma^2).$$ [4]

Ese supuesto conduce a dos resultados importantes para hacer inferencia estadística. El primero es que cada $\hat{\beta}_k$, $k = 1, \ldots, K$, también se distribuye normal y como es insesgada y con varianza $\sigma^2 (X'X)^{-1}_{ii}$, donde $(X'X)^{-1}_{ii}$ es el elemento *i*-ésimo en la diagonal de $(X'X)^{-1}$, se obtiene que

$$\hat{\beta}_k \sim N(\beta_k, \sigma^2\, (X'X)^{-1}_{ii})$$

y estandarizando se llega a

[4] Aquí solo se plantearan los intervalos de confianza en esos escenarios, para que se note la diferencia en las distribuciones de las cantidades pivotales. Las pruebas de hipótesis se efectuaran solo bajo el supuesto de que la varianza del término estocástico es desconocida, que comúnmente es lo que sucede cuando se hace un modelo.

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2\ (X'X)_{ii}^{-1}}} \sim N(0,1). \tag{3}$$

El otro resultado es que

$$(T-K)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-K}, \tag{4}$$

donde $\hat{\sigma}^2 = \hat{\sigma}^2_{MCO}$ y χ^2_{T-K} indica la distribución chi- cuadrada con *T*-*K* grados de libertad.[5] A partir de estas expresiones se formulan los intervalos de confianza y se realizan las pruebas de hipótesis.

3.2.1 Intervalos de confianza para $\hat{\beta}_k$

Partiendo de la expresión en (3) y suponiendo que σ^2 es conocida se llega después de un poco de álgebra, al intervalo de confianza

$$P\left(\hat{\beta}_k - z_{\alpha/2}\sqrt{\sigma^2\ (X'X)_{ii}^{-1}} \leq \beta_k \leq \hat{\beta}_k + z_{\alpha/2}\sqrt{\sigma^2\ (X'X)_{ii}^{-1}}\right) = 1-\alpha,$$

donde $z_{\alpha/2}$ es el valor crítico apropiado a una distribución normal y α es el nivel de significancia.

Cuando σ^2 es desconocida, se debe estimar y por ello, su estimador mínimo cuadrático se sustituye en (3) y entonces, la cantidad pivotal ya no se distribuye normal, tiene una distribución *t*-*S*tudent (*tS*) con *T*-*K* grados de libertad, es decir,[6]

---

[5] En este punto es importante recordar que la distribución χ^2 es el resultado de sumar el cuadrado de variables aleatorias independientes e idénticamente distribuidas de manera normal estándar, en el contexto del modelo de regresión, esas variables son los errores aleatorios, que se estandarizan para poder utilizar este resultado.

[6] Formalmente, esta cantidad pivotal y su distribución se obtiene mediante el cociente de la expresión con distribución normal en (3) y la raíz cuadrada de la χ^2 que se encuentra en (4) entre sus grados de libertad, esto es,

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2\ (X'X)^{-1}_{kk}}} \sim tS_{T-K}$$

y después de un poco de álgebra se obtiene el intervalo de confianza

$$P\left(\hat{\beta}_k - \tau_{\alpha/2}\sqrt{\hat{\sigma}^2\ (X'X)^{-1}_{ii}} \leq \beta_k \leq \hat{\beta}_k + \tau_{\alpha/2}\sqrt{\hat{\sigma}^2\ (X'X)^{-1}_{ii}}\right) = 1 - \alpha,$$

en él, $\tau_{\alpha/2}$ es el valor crítico asociado a una distribución *tS* con *T-K* grados de libertad.

3.2.2 Pruebas de hipótesis para $\hat{\beta}_k$[7]

Considere ahora que se quiere probar las siguientes hipótesis

$$H_0\text{: } \beta_k = b_k \quad \text{vs} \quad H_1\text{: } \beta_k \neq b_k,$$

donde b_k, $k = 1, \ . \ . \ . \ , K$, es una constante dada. Bajo el supuesto de que σ^2 no es conocida, la expresión en (3) aun cuando se sustituya el estimador de σ^2 no es un estadístico de prueba, puesto que el parámetro β_k es desconocido, pero bajo la hipótesis nula, β_k toma el valor b_k que sí se conoce, de manera que su sustitución en (3) conduce al siguiente estadístico de prueba bajo H_0,

$$\tau = \frac{\hat{\beta}_k - b_k}{\sqrt{\hat{\sigma}^2\ (X'X)^{-1}_{kk}}} \sim tS_{T-K}$$

$$\frac{\dfrac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2\ (X'X)^{-1}_{kk}}}}{\sqrt{\dfrac{(T-K)}{(T-K)}\dfrac{\hat{\sigma}^2}{\sigma^2}}} = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2\ (X'X)^{-1}_{kk}}}.$$

Como numerador y denominador son independientes, el cociente tiene una distribución *tS* cuyos grados de libertad son los de la χ^2. Recuérdese que el cociente de una distribución normal y la raíz cuadrada de una χ^2 dividida por sus grados de libertad, tiene una distribución *tS* que hereda los grados de libertad de la chi-cuadrada.

[7] Debido a que la varianza del término estocástico generalmente es desconocida, en lo que sigue se hacen las pruebas para el caso en el que no es conocida.

y la región crítica, donde se rechaza H_0 a un nivel de significancia α%, es

$$|\tau| > c_\alpha \tag{5}$$

donde c_α es el valor crítico asociado a α.

Otra forma equivalente, de determinar si la información empírica incorporada al modelo proporciona evidencia a favor o en contra de la hipótesis nula, es mediante el *p-value* o nivel de significancia marginal, se rechaza H_0 si y solo si

$$p\text{-}value < \alpha. \tag{6}$$ [8]

Dentro de estas pruebas de hipótesis tiene particular relevancia, la prueba significancia individual, es decir, la que asume bajo H_0 que $b_k = 0$, esto es,

$$H_0: \beta_k = 0 \quad \text{vs} \quad H_1: \beta_k \neq 0. \tag{7}$$

Esta es importante porque a través de ella se analiza si β_k es estadísticamente significativo, en cuyo caso, la variable que lo acompaña es importante desde el punto de vista estadístico, en la determinación de la variable dependiente.

3.2.3. Pruebas de hipótesis para combinaciones lineales de las betas

La prueba de hipótesis asociada a combinaciones lineales de los parámetros es una prueba conjunta que al igual que los intervalos de confianza y las pruebas de hipótesis estudiadas con anterioridad, basa su desarrollo en la normalidad del término estocástico del modelo de regresión. Considerando que se tienen *m* combinaciones lineales de los parámetros beta, las hipótesis a probar son

$$H_0: R\beta = r \quad \text{vs} \quad H_0: R\beta \neq r,$$

[8] Un análisis detallado de estos aspectos se encuentran en Davidson y MacKinnon (2004) y Spanos (1999).

donde R es la matriz de los coeficientes de las combinaciones lineales y es de dimensión $m \times K$, β es el vector que contiene los K parámetros beta y r es un vector de dimensión m con los términos independientes de cada restricción o combinación lineal. Ejemplos de este tipo de pruebas se encuentran en Johnston y Dinardo (1997), así como un estudio exhaustivo de las mismas.

La normalidad del error estocástico conduce también a que

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$$

y por tanto,

$$R\hat{\beta} \sim N(R\beta, \sigma^2 R(X'X)^{-1}R'),$$

lo cual conduce a

$$R\hat{\beta} - R\beta \sim N(\mathbf{0}, \sigma^2 R(X'X)^{-1}R')$$

donde el vector $\mathbf{0}$ es de dimensión K y finalmente,

$$\left(R\hat{\beta} - R\beta\right)' [\sigma^2 R(X'X)^{-1}R']^{-1} (R\hat{\beta} - R\beta) \sim \chi^2_{m.}$$

Esta expresión y la planteada en (4) son formas cuadráticas independientes, cuyo cociente dividido numerador y denominador por sus correspondientes grados de libertad, conduce al estadístico de prueba que se muestra a continuación y que bajo H_0, se distribuye como una F con m y T-K grados de libertad,

$$f = \frac{1}{m\hat{\sigma}^2} \left(R\hat{\beta} - r\right)' [R(X'X)^{-1}R']^{-1} \left(R\hat{\beta} - r\right) \sim F_{(m, T-K)}$$

Los criterios de rechazo de la hipótesis nula son igual a los planteados en (5) y (6) usando el valor crítico o el *p-value*, pero ahora se debe usar la distribución *F* para decidir si rechazar o no H_0.[9]

Aquí es importante señalar dos aspectos dentro de este tipo de pruebas. El primero es que al igual que la de significancia individual, es decir, la que considera un sólo parámetro y que fue formulada en (7), la de significancia conjunta también tiene relevancia en la evaluación econométrica del modelo, en ella se considera bajo la hipótesis nula que los *K*-1 parámetros que son coeficientes de las variables independientes, son cero, de manera que las hipótesis se plantean como

$$H_0: R\beta = \mathbf{0} \quad \text{vs} \quad H_0: R\beta \neq \mathbf{0},$$

donde

$$R = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

y $\mathbf{0}$ es un vector de ceros de dimensión *m*=*K*-1, o equivalentemente,

$$H_0: \beta_k = 0, \forall k = 2, \ldots., K$$

vs

$$H_1: \beta_k \neq 0, \text{para al menos una } k = 2, \ldots, K.$$

De esta manera, en caso de que se rechace H_0, habrá evidencia a favor de que las variables independientes del modelo son estadísticamente diferentes de cero y por tanto, son relevantes estadísticamente para explicar a la variable independiente.

[9] Los criterios para el rechazo o no de la hipótesis nula siempre son los mismos, pero se debe saber tanto la distribución apropiada del estadístico de prueba para determinar el valor crítico y el *p-value*, como la hipótesis nula.

El otro punto a resaltar es que esta es la prueba más general y por tanto, la de significancia individual es un caso particular, en ella la matriz R es de la siguiente forma

$$R = (0 \quad 0 \quad \cdots \quad 1 \quad \cdots \quad 0 \quad 0)$$

es de dimensión $K \times 1$ y el uno está en el lugar k-ésimo de R.

3.2.4 Intervalo de confianza y prueba de hipótesis para σ^2

A partir de la expresión en (4) dada por

$$(T-K)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-K}, \tag{8}$$

se puede plantear después de un poco de álgebra, el siguiente intervalo de confianza para la varianza del error estocástico, σ^2,

$$P\left((T-K)\frac{\hat{\sigma}^2}{\chi_{\alpha/2}} \leq \sigma^2 \leq (T-K)\frac{\hat{\sigma}^2}{\chi_{1-\alpha/2}}\right) = 1-\alpha,$$

y también se pueden probar las hipótesis

$$H_0: \sigma^2 = s \quad \text{vs} \quad H_1: \sigma^2 \neq s,$$

donde $s > 0$ es una constante conocida y por tanto, bajo la hipótesis nula el estadístico de prueba es

$$(T-K)\frac{\hat{\sigma}^2}{s^2} \sim \chi^2_{T-K},$$

que resulta de sustituir en (8) el valor de σ^2 bajo la hipótesis nula.

4. Prueba de normalidad de Jarque-Bera

Cuando una variable aleatoria se distribuye normal, su tercer y cuarto momento alrededor de la media también conocidos como sesgo y curtosis, son cero y tres, de manera respectiva. El sesgo igual a cero da cuenta de que la distribución es simétrica, mientras que la curtosis igual a tres plantea que la distribución no es puntiaguda (leptocúrtica), ni achatada (platicúrtica), en cuyo caso es normal o mesocúrtica.

Jarque y Bera (1980, 1987) formulan una prueba de normalidad que lleva su nombre, ellos plantean que existen distribuciones que pueden coincidir con la distribución normal, en media y varianza o sea, que su primer momento centrado en cero y su segundo alrededor de la media son los mismos, pero que no necesariamente el tercero y cuarto momentos centrados en la media son iguales. Esa es la razón que los conduce a plantear la prueba de normalidad basada en el sesgo, *s*, y la curtosis, *c*, de manera que las hipótesis a probar son

$$H_0\text{: Errores normales} \quad \text{vs} \quad H_1\text{: Errores no normales}$$

o equivalentemente,

$$H_0: s = 0, c = 3 \quad \text{vs} \quad H_1: s \neq 0 \ \text{y/o} \ c \neq 3$$

y el estadístico de prueba bajo H_0 es

$$JB = T\left[\frac{\widehat{cs}^2}{6} + \frac{(\widehat{cc} - 3)^2}{24}\right] \sim \chi^2_{(2)}$$

donde $\widehat{cs}$ es el coeficiente de sesgo y el $\widehat{cc}$ coeficiente de curtosis dados por

$$\widehat{cs} = \frac{\hat{s}}{\left(\sqrt{\hat{\sigma}^2}\right)^3}, \qquad \widehat{cc} = \frac{\hat{c}}{\left(\sqrt{\hat{\sigma}^2}\right)^4}$$

y

$$\hat{s} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^3, \qquad \hat{c} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^4.$$

Observe que si el estadístico de prueba *JB* es cercano a cero hay evidencia a favor de que los errores se distribuyen de manera normal, en caso contrario, es decir, cuando *JB* está alejado de cero, se rechaza la hipótesis nula y las distribuciones de los estimadores de las betas y de la varianza de los errores estocásticos, no tendrán las distribuciones que permiten hacer inferencia estadística del modelo de regresión lineal y por tanto, ellas serán válidas sólo de manera asintótica de acuerdo al teorema de límite central.

5. Prueba Jarque-Bera en R

En la implementación de la prueba de Jarque-Bera en R, es necesario cargar el paquete tseries mediante la instrucción

```
> library(tseries)
```

y una vez que se cuenta en el objeto al que se la va aplicar la prueba se debe introducir

```
> jarque.bera.test(x)
```

en donde el argumento x es un vector o una serie de tiempo. Esta prueba puede llevarse a cabo para cualquier vector o serie de tiempo del que se desee saber si se distribuye o no de manera normal. Sin embargo, en el contexto del modelo de regresión, la prueba se realiza sobre los residuales ya que estos son las *proxis* de los errores estocásticos que se suponen son normales, por ello la instrucción para efectuar la prueba de Jarque-Bera para los residuales del objeto llamado model, que guarda los resultados de estimación de la regresión, es

```
> jarque.bera.test(residuals(model))
```

y obviamente, debe de ejecutarse una vez que se estima el modelo. El resultado de la prueba presenta el nombre de la variable en `data`, el estadístico de prueba en `X-squared`, los grados de libertad en `df` y el mínimo nivel de significancia al que se rechaza la hipótesis nula, en `p-value`.

Ejemplo 1.

En este ejemplo se genera una variable (vector), que contiene cien números seleccionados de manera aleatoria de una distribución normal y se efectúa la prueba de normalidad para ese variable, pero antes se instala el paquete tseries, tal y como se muestra a continuación.

```
> library(tseries)

    'tseries' version: 0.10-34
```

```
    'tseries' is a package for time series analysis and
    computational finance.

    See 'library(help="tseries")' for details.

> y<-rnorm(100)
> jarque.bera.test(y)

        Jarque Bera Test

data:  y
X-squared = 0.46901, df = 2, p-value = 0.791
```

A un nivel de significancia del 5%, la hipótesis nula de normalidad no es rechazada, puesto que *p-value*>0.05.

Ejemplo 2.

La información anual de 1953 a 2004 contenida en el archivo Gasolina.txt es usada para estimar un modelo para la demanda de gasolina en USA (Greene, 2003). Se plantean dos regresiones log-log, en la primera se modela la demanda per-cápita en función del ingreso per-cápita, del índice de precios de la gasolina y el de los autos nuevos y en la segunda, se agrega el índice de precios agregado del consumo de bienes durables, y en ambas se prueba normalidad.[10] Las instrucciones en R son las que se presentan a continuación.

```
> library(tseries)

    'tseries' version: 0.10-34

    'tseries' is a package for time series analysis and
    computational finance.

    See 'library(help="tseries")' for details.

> Gasolina <- read.csv("Gasolina.txt")
>   View(Gasolina)
> attach(Gasolina)
```

[10] Las variables del archivo son

Año: 1953-2004,
G: Gasto total en gasolina,
Pobl: Población
Pg: Indice de precio de la gasolina,
Y: Ingreso disponible per-cápita,
Pan: Indice de precios de los autos nuevos,
Pau: Indice de precios de los autos usados,
Ptp: Indice de precios del transporte público,
Pd: Indice de precios agregado del consumo de bienes durables,
Pnd: Indice de precios agregado del consumo de bienes no durables,
Ps: Indice de precios agregado para el consumo de servicios.

Fuente: http://people.stern.nyu.edu/wgreene/Text/econometricanalysis.htm

```
> cons<-lm(log(G/Pobl)~log(Y)+log(Pg)+log(Pan))
> jarque.bera.test(residuals(cons))

        Jarque Bera Test

data:  residuals(cons)
X-squared = 7.3104, df = 2, p-value = 0.02586

> cons<-lm(log(G/Pobl)~log(Y)+log(Pg)+log(Pan)+log(Pd))
> jarque.bera.test(residuals(cons))

        Jarque Bera Test

data:  residuals(cons)
X-squared = 3.6263, df = 2, p-value = 0.1631
```

Con base en los resultados de la prueba de Jarque-Bera de la primera regresión, se rechaza la hipótesis nula de normalidad al 5%, puesto que *p-value*<0.05, pero no al 1% de significancia (*p-value*>0.01), mientras que en la segunda regresión no se rechaza la hipótesis nula y por ello se infiere que los errores son normales.

Ejemplo 3

En este ejemplo se presenta un modelo estático para la elasticidad de sustitución Armington para México.[11] La estimación incorpora información trimestral del INEGI para el periodo que comprende del primer trimestre de 1993 al primero del 2013, a precios constantes del 2008. La variable explicada en el modelo es la demanda relativa (DRel), que resulta del cociente entre las importaciones totales y la demanda doméstica (diferencia entre el valor bruto de la producción y las exportaciones, ambas a precios de mercado) y las variables explicativas son el precio relativo (PRel), que se obtiene de dividir el índice de precios de la demanda relativa entre el correspondiente a las importaciones, y el producto interno bruto (PIB).

Las siguientes instrucciones en R permiten estimar el modelo de regresión log-log con las variables descritas previamente y efectuar la prueba de normalidad de Jarque-Bera.

```
> Elast <- read.csv("Elast.txt")
>   View(Elast)
> attach(Elast)
> model<-lm(log(DRel)~log(PRel)+log(PIB))

> jarque.bera.test(residuals(model))
```

[11] Un análisis detallado de este modelo desde sus microfundamentos hasta la especificación final de un modelo dinámico es presentado en Casares, Ruiz-Galindo y Sobarzo (por publicarse).

```
        Jarque Bera Test

data:  residuals(model)
X-squared = 4.9739, df = 2, p-value = 0.08316
```

El *p-value* implica que la hipótesis nula de normalidad de los errores aleatorios no se rechaza a un nivel de significancia del 5%, pero si al 10%, puesto que *p-value*>0.05 y *p-value*<0.10.

6. Causas e implicaciones de la no normalidad y posibles soluciones

Dos son las causas principales de que los residuos del modelo no se distribuyan de manera normal: una es que la muestra no es lo suficientemente grande como para garantizarla y la otra es que si a los datos que se incorporaron al modelo se les hizo alguna transformación, ella no fue la adecuada.

Cuando los datos son pocos y hay posibilidad de obtener más, habrá que incluirlos para obtener una nueva estimación del modelo. Si esto no es posible, habrá que hacer una transformación de la familia Box y Cox, de las cuales la más utilizada es la logarítmica, y que además también puede corregir heteroscedasticidad.

Considere que se quiere transformar a variable *w* cuyos valores son positivos, la transformación Box-Cox depende de un parámetro λ y es la siguiente

$$z(\lambda) = \begin{cases} \frac{w^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \log w, & \lambda = 0 \end{cases}.$$

Cuando *w* no es positiva, se le suma una constante de manera que se obtengan una nueva variable cuyos valores sí lo sean.

7. Conclusiones

La elaboración de los modelos econométricos conlleva dos tipos de evaluación. Una que se basa en la teoría económica que fue utilizada para la especificación del mismo y en la que se revisan que los signos de los parámetros estimados y su magnitud, entre otros aspectos, coincidan con los que formula la teoría. La otra es la evaluación econométrica, que consiste en analizar la significancia individual y conjunta de los parámetros y verificar sí se satisfacen tanto los supuestos del modelo de regresión en su parte determinista como los que se plantean en los términos o errores aleatorios.

Dentro de la evaluación econométrica reviste importancia la normalidad, ya que este supuesto aunque no necesario en la estimación de los parámetros del modelo, resulta indispensable en las otras dos formas de hacer inferencia estadística, a saber, en el planteamiento de los intervalos de confianza y de las pruebas de hipótesis. A partir de la normalidad de los errores aleatorios, se obtiene las distribuciones apropiadas de las cantidades pivotales para plantear intervalos de confianza y de los estadísticos de prueba para efectuar pruebas de hipótesis.

Por lo anterior, una vez que se ha analizado que los parámetros estimados tienen los signos y magnitudes apropiadas de acuerdo a la teoría económica subyacente, se debe revisar si los residuos del modelo que son las *proxis* de los términos estocásticos, son normales, de no ser así se corre el riesgo de hacer inferencia de manera incorrecta a menos que se tenga una gran cantidad de observaciones para cada variable, en cuyo caso se recurre al teorema de límite central que garantiza normalidad cuando el tamaño de la muestra tiende a infinito, en la práctica esto significa que se tienen muchas observaciones y por tanto, los resultados de inferencia estadística expuestos aquí son válidos de manera asintótica. Sin embargo, si no se puede incrementar el número de observaciones o bien a pesar de haberlo hecho no se obtuvo normalidad, se debe usar una transformación de las variables de la familia Box-Cox y de manera específica, la logarítmica que es la más utilizada en estas situaciones.

Bibliografía

Casares, E. R., L. A. Ruiz-Galindo y H. Sobarzo, (por publicarse). “Short and Long Run Armington Elasticities for the Mexican Economy” en A. Pinto y D. Zilberman (editors), Modeling, Dynamics, Optimization and Bioeconomics II, en la serie Springer Proceedings in Mathematics an Statistics.

Davidson R. y J. G. MacKinnon, (2004). Econometric Theory an Methods. Ed. Oxford University Press, New York.

Greene, W. H., (2007). Econometric Analysis. Ed. New York University, New York.

Jarque, C. M. y A. K. Bera (1980). "Efficint test s for normality, heteroskedasticity and serial independdence of regression residuals", *Economics Letters*, vol. 6, 255-259.

Jarque, C. M. y A. K. Bera (1987). "A Test for Normality of Oservations and Regression Residuals", *International Statistical Review*, vol 55, 2,163-172.

Johnston, J. y J. Dinardo, (1997). *Economerics Methods* Ed. McGraw-Hill, Singapur.

Spanos, A., (1999). *Probability Theory and Statistical Inference. Econometric Modeling with Observational Data*. Ed. Cambridge University, Reino Unido.

Referencias electrónicas

Datos (Greene, 2007), http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm

INEGI (2013a), "Banco de Información Económica", http://dgcnesyp.inegi.gob.mx

Archivos de las bases de datos de este Capítulo

Gasolina.txt

Elast.txt

ECONOMETRÍA
APLICADA UTILIZANDO R.

PAPIME PE302513 LIBRO ELECTRÓNICO Y COMPLEMENTOS DIDÁCTICOS EN MEDIOS COMPUTACIONALES, PARA EL FORTALECIMIENTO DE LA ENSEÑANZA DE LA ECONOMETRÍA

CAPÍTULO 5
Normalidad

Lucía A. Ruiz Galindo*

* Departamento Economía, UAM-A

FES ACATLÁN
dgapa
SAREC

Normalidad

Modelo General de Regresión Lineal

- Especificación

$$y_t = \beta_1 + \beta_2 x_{t2} + \ldots + \beta_K x_{tK} + \varepsilon_t,$$

donde $\beta_1, \ldots, \beta_K$ son los parámetros del modelo,

y_t es la variable dependiente,

x_{tk}'s, $k = 2, \ldots, K$, son las variables independientes,

ε_t es el término o error estocástico,

t, $t = 1, \ldots, T$, es un índice que indica el número de la observación y

T es el total de observaciones

Normalidad

Modelo General de Regresión Lineal

- Supuestos en la forma funcional

S1. Linealidad en los parámetros.

S2. Las K-1 variables independientes son las únicas que explican a la dependiente.

S3. El número de observaciones T, es mucho mayor que el de parámetros K.

S4. Las variables explicativas son linealmente independientes de manera que ninguna es combinación lineal de otra o de otras y por tanto el rango de X es K.

S5. Los parámetros no cambian en la muestra, es decir, hay permanencia estructural

Normalidad

Modelo General de Regresión Lineal

- Supuestos Gauss-Markov

SGM1. $E(\varepsilon_t) = 0,\ \forall t = 1,\ \ldots, T.$

SGM2. $\{\varepsilon_1, \varepsilon_2,\ \ldots, \varepsilon_T\}$ y $\{x_{1k}, x_{2k},\ \ldots, x_{Tk}\}$ son independientes $\forall k = 2,\ \ldots, K.$

SGM3. $V(\varepsilon_t) = \sigma^2,\ \forall t = 1,\ \ldots, T.$

SGM4. $Cov(\varepsilon_t, \varepsilon_s) = 0,\ \forall t, s = 1,\ \ldots, T,\ t \neq s.$

SGM5. ε_t se distribuye Normal, $\forall t = 1,\ \ldots, T.$

Normalidad

Importancia Normalidad en la Inferencia Estadística

- Estimación Puntual: MCO y MV

MV: $\varepsilon_1, \varepsilon_2, \ . \ . \ . \ , \varepsilon_T$ constituyen una muestra aleatoria

Función de verosimilitud y log-verosimilitud

$$L(\beta_1, . \ .. , \beta_K; \ \varepsilon_1, . \ . \ . \ , \varepsilon_T) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{T}{2}} e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{T} \varepsilon_t^2}$$

$$l(\beta_1, . \ . \ . \ , \beta_K; \ \varepsilon_1, . \ . \ . \ , \varepsilon_T) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \varepsilon_t^2$$

Normalidad

Importancia Normalidad en la Inferencia Estadística

- Distribuciones de los parámetros

  - $\hat{\beta}_k \sim N\left(\beta_k, \sigma^2 \, (X'X)_{ii}^{-1}\right)$

  - $(T-K)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-K}$

Normalidad

Importancia Normalidad en la Inferencia Estadística

- Intervalos de confianza para β_k.
  - σ^2 conocida, $\dfrac{\widehat{\beta}_k - \beta_k}{\sqrt{\sigma^2\ (X'X)_{ii}^{-1}}} \sim N(0,1)$
  - σ^2 desconocida, $\dfrac{\widehat{\beta}_k - \beta_k}{\sqrt{\widehat{\sigma}^2\ (X'X)_{kk}^{-1}}} \sim tS_{T-K}$.

$$P\left(\hat{\beta}_k - \tau_{\alpha/2}\sqrt{\hat{\sigma}^2\ (X'X)_{ii}^{-1}} \leq \beta_k \leq \hat{\beta}_k + \tau_{\alpha/2}\sqrt{\hat{\sigma}^2\ (X'X)_{ii}^{-1}}\right) = 1 - \alpha$$

Normalidad

Importancia Normalidad en la Inferencia Estadística

- Prueba de hipótesis para β_k (σ^2 desconocida).

$$H_0\text{: } \beta_k = b_k \quad \text{vs} \quad H_1\text{: } \beta_k \neq b_k$$

Estadístico de prueba bajo H_0:

$$\tau = \frac{\hat{\beta}_k - b_k}{\sqrt{\hat{\sigma}^2 \, (X'X)^{-1}_{kk}}} \sim tS_{T-K}$$

Normalidad

Importancia Normalidad en la Inferencia Estadística

- Prueba de hipótesis para combinaciones lineales de β_k´s (σ^2 desconocida).

$$H_0:\ R\beta = r \quad \text{vs} \quad H_0:\ R\beta \neq\ r$$

Estadístico de prueba bajo H_0:

$$f = \frac{1}{m\hat{\sigma}^2}\left(R\hat{\beta} - r\right)'[R(X'X)^{-1}R']^{-1}\left(R\hat{\beta} - r\right) \sim F_{(m,T-K)}$$

Normalidad

Importancia Normalidad en la Inferencia Estadística

- Prueba de significancia estadística

  - Individual

$$H_0\colon\ \beta_k = 0 \quad \text{vs} \quad H_1\colon\ \beta_k \neq 0$$

  - Conjunta

$$H_0\colon\ R\beta = 0 \quad \text{vs} \quad H_0\colon\ R\beta \neq 0$$

Normalidad

Importancia Normalidad en la Inferencia Estadística

- Intervalo de confianza para σ^2

$$(T-K)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-K}$$

$$P\left((T-K)\frac{\hat{\sigma}^2}{\chi_{\alpha/2}} \leq \sigma^2 \leq (T-K)\frac{\hat{\sigma}^2}{\chi_{1-\alpha/2}}\right) = 1-\alpha$$

Normalidad

Importancia Normalidad en la Inferencia Estadística

- Prueba de hipótesis para σ^2

$$H_0\colon \sigma^2 = s \quad \text{vs} \quad H_1\colon \sigma^2 \neq s,$$

Estadístico de prueba bajo H_0:

$$(T-K)\frac{\hat{\sigma}^2}{s^2} \sim \chi^2_{T-K}$$

Normalidad

Prueba de Normalidad de Jarque-Bera (JB)

- Prueba estadística

$$H_0\colon\ s = 0, c = 3 \quad \text{vs} \quad H_1\colon\ s \neq 0 \ \text{y/o} \ c \neq 3$$

- Estadístico de prueba

$$JB = T\left[\frac{\widehat{cs}^2}{6} + \frac{(\widehat{cc} - 3)^2}{24}\right] \sim \chi^2_{(2)}$$

Normalidad

Prueba de Jarque-Bera en R

- Cargar paquete tseries

```
>library(tseries)
```

- Disponer del objeto

- Aplicar la prueba

```
>jarque.bera.test(x)

>jarque.bera.test(residuals(model))
```

Normalidad

Ejemplos de la Prueba de Jarque-Bera en R

- Ruido blanco

```
>  library(tseries)
'tseries' version: 0.10-34
'tseries' is a package for time series analysis and
computational finance.
See 'library(help="tseries")' for details.
>  y<-rnorm(100)
>  jarque.bera.test(y)
            Jarque Bera Test
data:  y
X-squared = 0.46901, df = 2, p-value = 0.791
```

Normalidad

Ejemplos de la Prueba de Jarque-Bera en R

- Elasticidades Argmington

```
> Elast <- read.csv("Elast.txt")
>     View(Elast)
> attach(Elast)
> model<-lm(log(DRel)~log(PRel)+log(PIB))

> jarque.bera.test(residuals(model))

        Jarque Bera Test

data:  residuals(model)
X-squared = 4.9739, df = 2, p-value = 0.08316
```